12

6 DATACOMPUTER SUPPORT OF SEISMIC DATA ACTIVITY.

9 Quarterly Technical Report.
1 Aug—31 Oct 76.

11 31 Oct 76

12 13p.

August 1, 1976 to October 31, 1976

15 Contract No. MDA903-74-C-0227
ARPA Order No. 2613

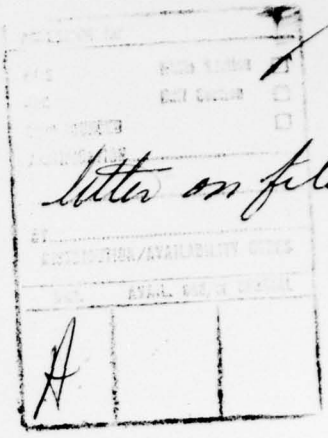
Submitted to:

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: John Filson

387285

B



Computer Corporation of America
575 Technology Square
Cambridge, Massachusetts 02139

Datacomputer Support of Seismic Data Activity

Quarterly Technical Report

August 1, 1976 to October 31, 1976

This research is supported by the Defense Advanced Research Projects Agency of the Department of Defense under Contract No. MDA903-74-C-0227. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

Table of Contents

1. Introduction

The purpose of this project is to provide seismic data storage and retrieval services. These services are provided via the Arpanet Datacomputer being developed and maintained by Computer Corporation of America (CCA) for ARPA under a separate Contract No. MDA903-74-C-0225. The seismic data is to received over the Arpanet and made available to computers on the Arpanet in a convenient and timely manner.

To provide the requisite services the Datacomputer has been augmented by a mass memory system to provide the very large storage resources required. To deal with the high seismic data rates CCA has implemented a Seismic Input Processor (SIP) which continuously collects real time data over the network, reformats and buffers it, and periodically bursts it into the Datacomputer. An Ampex Tera-Bit Memory System (TBM) with a capacity of almost two hundred billion bits (expandable to 3 trillion bits) has been installed at CCA as the mass memory system. A DEC 11/40 computer with RP-04 disks and an Arpanet interface has been purchased as the SIP. Both the SIP and the Datacomputer with mass memory are operational.

Project activity can be divided into four catagories: (1) integration of the mass memory into Datacomputer and other seismic related Datacomputer development; (2) the acceptance and use of the mass memory; (3) coordination with the seismic community; and (4) the SIP and network bandwidth considerations. One section in this report is devoted to each of these categories.

2. The Datacomputer

The Datacomputer is a network data utility developed by CCA. It is designed to handle large files and to communicate with multiple remote using programs over the Arpanet. The Datacomputer is the primary on-line repository for seismic data.

2.1 Version 2

The Version 2 Datacomputer is the first generally available version of the system which uses a mass memory (Ampex TBM) as its tertiary store. An experimental Datacomputer with mass memory (called the 103 Datacomputer) was available to a limited set of users in August and September. It was terminated on 19 October. Use of the Version 2 Datacomputer began on 1 October. The Version 2 Datacomputer incorporated changes that were found necessary through experience with the 103. Among the most important of these are the following:

1. Volume interlocks, to coordinate two users who are trying to access several blocks at two distant spots on a tape. Without interlocks such uses can cause the drive to spend most of its time seeking back and forth between the two spots alternately accessing one block in each area. In Version 2 each request will try to seize a lock associated with the drive and, if it succeeds, will proceed until it reaches a natural interruption point.

2. Special subforks under the Datacomputer subjobs. This enhancement was made to increase the virtual memory space available; this was necessary due to the large size of TBM blocks (over a million bits) and the correspondingly large buffers they require.

3. A general file backup facility. This feature is used to automatically copy most permanent files for redundancy. It is used either automatically or manually to copy parts of files that are on worn areas of TBM tape.

2.2 Version 3

Work continues toward the next version of the Datacomputer, Version 3, which is scheduled to be released at the end of calendar 1976.

The most significant enhancements for Version 3 are better accounting facilities and the file groups feature.

File groups will enable several similar physical files to be accessed as though they were a single file. This is particularly useful for seismic data streams since, by nature they grow indefinitely over time. However for convenience in allocation, creation, dumping, validating, access control, and ultimately due to the finite size of TBM tapes, these continuous streams must be divided into finite physical files. The file group feature allows a user to ignore these physical boundaries on retrieval.

The accounting facilities will record the resources used from each login node and will aggregate these to show the total resources attributable to special nodes called "billable" nodes. This feature will permit easier supervision of the resource utilization by members of the seismic community.

2.3 TENEX

During this quarter the Datacomputer operating system, TENEX, was modified by CCA to improve its Arpanet and TBM interfaces.

The Arpanet interface was modified by:

1. increasing the number of buffers available,
2. increasing the number of network conncections which can exist simultaneously, and
3. decreasing the size of the M3 network message padding field.

The TBM interface was improved by:

1. Adding a system call to activate the Auto Align facility (see 3.3), and
2. increasing the fault tolerance of TENEX with respect to TBM errors.

It is expected that further modifications to TBM error handling will be advisable as CCA learns more about its actual performance and failure modes.

3. The Mass Memory Subsystem

To provide on-line storage approaching the large amount required, a mass memory, the Ampex Tera-Bit Memory System (TBM), has been integrated into the Datacomputer. The TBM has four tape drives each with a capacity of nearly 50 billion bits for a total maximum of near 200 billion bits on line. The drives are physically similar to professional quality videotape drives.

Each reel of TBM tape has a unique number associated with it and is pre-formatted into fixed numbered blocks. 43,800 of these blocks per tape are user accessible. Other blocks are reserved for internal TBM use. Besides its number, each block has associated with it, in a separately recorded "tally track", a file data identification number which is used for block address error checking and various other information including counts of operations performed to the block.

3.1 Acceptance

The TBM hardware was accepted last quarter but funds were withheld pending correction of problems in the basic internal TBM software. A supplementary software acceptance test was completed by early September. The only outstanding items due to CCA from Ampex are the Auto Align and Read Recovery enhancements (see 3.4) and completion of TBM operator training for CCA personnel.

3.2 Performance

Considering the short period that the TBM has been in operation, performance has been generally satisfactory both mechanically and electronically. Some electronic design flaws, as described below, were discovered and fixed. The only serious problem has

been an infrequent garbling of the "tally track" information associated with some blocks on TBM tape. These errors make it impossible to read or write effected blocks without manual intervention. CCA is working with Ampex to isolate and solve this tally track problem.

Three electronic design problems have been discovered in the TBM, all of the same general nature. Some of the circuitry in the TBM used logic elements only marginally fast enough for their function. As a result, the normally negligible effects of several months aging and use caused these circuits to fail intermittantly. These problems have been fixed by replacing the logic with compatible higher speed elements.

3.3 Enhancements

Ampex is required to install two enhancements to the TBM, known as Auto Align and Read Recovery, which are necessary to convenient operation of the TBM as a component of the Datacomputer. The necessary changes to the TBM internal software are discribed below. Initial installation and testing of these enhancements began during the reporting period. A 7½% retainage on the price of the TBM is being withheld pending completion of these features.

In order to compensate for rotary head wear and other gradual changes in the characteristics of a TBM drive, tape, and the system as a whole, it is periodically necessary to "align" a drive with the tape one desires to use. The alignment process yields a table of parameters that are remembered by the TBM system indexed under the drive and tape numbers. Alignment is currently a manual operation which can take up to 15 minutes and which requires taking the TBM system off line. The Auto Align enhancement will enable the TBM to go through an alignment sequence automatically. The automatic alignment process requires not more than two minutes to complete. The TBM will remain on-line with other operations queued during the alignment as they would be during a two minute sequence of data operations.

The read recovery feature facilitates operations when a read error is encountered. In the case of difficulty in reading a block from TBM tape, there are a number of manual actions that can be taken at the TBM. These include disabling tally track processing and modifying the alignment parameters. Manually stepping through these actions is a tedious and error prone process which requires a skilled human operator. The Read Recovery enhancement will enable the system to try these recovery actions automatically. The Read Recovery operation will stop as soon as it is successful or will fail after trying all plausible individual actions. The complete sequence requires not more than three minutes to complete.

3.4 Utilities

CCA has developed several utility programs for testing the TBM system. One of these utilities tests the proper operation of a drive. This checking process takes about five minutes and was previously uninterruptable. Other TBM users were denied access during this period. In October, this utility was modified so that its TBM use was interruptable at several points. This increases the time it takes for the checker to run if it is in active conflict with a Datacomputer user but avoids locking out other TBM access. The checker is normally run daily on all drives after TBM maintenance, and on each drive when a new tape is mounted or when a high error rate is being encountered. As our experience with the TBM grows, we will continue to adapt and develop utility programs to facilitate its operation as a Datacomputer component.

4. Coordination with the Seismic Community

CCA has assisted VSC (Vela Seismological Center), SDAC (Seismic Data Analysis Center), ASL (Albuquerque Seismological Laboratory) and its contractors, and LL-ASG (Lincoln Laboratories Applied Seismology Group), in utlizing the Datacomputer to store and retrieve seismic data.

At the request of VSC, CCA suggested a full set of files and ports

for the non-array seismic data that is to be stored into the Datacomputer by ASL.

To provide a mechanism for announcing the state of the Datacomputer, status files were set up in the CCA directory at ISI and programs were written for updating these files.

At the request of various seismic users, some enhancements were made to RDC. RDC is a user program that runs under TENEX and simply interfaces a terminal to the Datacomputer. The Datacomputer is designed to be used by remote programs and cannot be used directly by a terminal.

5. The SIP and Arpanet Considerations

Seismic array data from LASA and NORSAR passes through the SDAC Communication and Control Processor (CCP) and from there is forwarded in real time to several locations including the SIP. The SIP buffers and reformats this data and periodically retransmits it to the Datacomputer. (non-real time seismic data is sent directly to the Datacomputer.) The SIP is equiped with disk storage adequate for twenty four hours buffering of a 15 kilobit per second data stream.

During this quarter the SIP became fully operational. After previous experiments with small test files, it forwarded seismic data into full size final files in a developmental TBM Datacomputer for August and September. At the beginning of October, the SIP output was switched to the Version 2 Datacomputer.

The real time data stream from the CCP to the SIP, together with high speed bursts from the SIP to the Datacomputer and other seismic and non-seismic Datacomputer traffic, all flowing over the Arpanet, have lead to communications overloads as detailed in our Quarterly Technical Reports for May 1, 1976 to July 31, 1976, and February 1, 1976 to April 30, 1976. This quarter,

the Lincoln Laboratories VDH line to the CCA IMP was moved to another IMP. This freed up enough resources at CCA IMP to make network performance adequate at current load levels. However, predictable growth in network traffic will cause increasingly severe problems. The only long-term solution appears to be the installation of a higher capacity IMP at CCA.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Computer Corporation of America	Unclassified
	2b. GROUP

3. REPORT TITLE

DATACOMPUTER SUPPORT OF SEISMIC DATA ACTIVITY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Quarterly Technical Report - August 1, 1976 to October 31, 1976

5. AUTHOR(S) (First name, middle initial, last name)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
MDA903-74-C-0227	
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Defense Advanced Research Projects 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Datacomputer						
Seismic Data Base						
Seismic Input Processor (SIP)						
Ampex TBM						